Appendix A:

Technical Notes

This appendix has discussions of the following topics in alphabetical order:

Age-adjusted rates
Attributable risk
Causes of death: ICD-9 codes
Causes of early infant death
Classification of data
Confidence intervals
Confidentiality
Congenital anomalies
Crude rates
Data quality
Event count
Hispanic origin
Infant mortality rates
Low birth weight
PUMS
Population data (denominators)
Prenatal care considerations
Race and ethnicity
Small numbers

Age-adjusted rates

Sometimes population characteristics need to be considered in interpreting comparisons. For example, since death rates increase with increasing age, a county with an older population may have higher death rates just because its population is older.

To compare rates and see if the county is high just because of its older population, we need to use age-adjusted death rates. These are computed by taking a county's death rates for each age group and applying them to a standard population (generally the 1940 U.S. population). This calculation then tells us what the county's death rate would be if it had the same age distribution as the U.S. population did in 1940. The major use of age-adjusted death rates is to allow comparisons between different geographic areas and/or over various periods of time.

Most death rates in this publication are age-adjusted to the 1940 U.S. population, which facilitates comparison with national rates and goals in the national publication *Healthy People*

2000, which also uses the 1940 standard. The exception to this rule is cancer death rates, which are age-adjusted to the 1970 U.S. population to facilitate comparison with statistics issued by the National Cancer Institute.

Users should be aware that an age-adjusted death rate has no absolute meaning; it is an artificial number based on a hypothetical population and is only useful for comparing with other rates calculated in the same manner. For more discussion of age-adjusted death rates, see the report of the Washington Department of Health, Center for Health Statistics, *Washington State Age-Adjusted Death Rates and Years of Life Lost Rates* (April 1992). While age adjustment is the most common method for adjusting rates, a similar process can be used to adjust for other characteristics such as sex, education, or birth weight.

Attributable risk

The percent of disease that could be eliminated by eliminating a known cause or risk factor.

Causes of Death: ICD-9 Codes

The causes of death presented in this document are classified in accordance with the International Classification of Diseases, Ninth Revision published by the World Health Organization. Washington state began using this revision on January 1, 1979.

Causes of early infant death

Causes of early infant death include the following ICD-9 codes: Maternal Diseases Affecting Child (760), Complications of Pregnancy Affecting Child (761-762), Birth Injuries (763, 767-769), Other Diseases Peculiar to Early Infancy (779).

Classification of Data

Classification and coding of data on Washington State vital records follow National Center for Health Statistics (NCHS) guidelines

as defined in 'Vital Statistics Instruction Manuals' parts 1-20 (Published by US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville MD).

Confidence intervals

Confidence intervals are presented in this document only for rates based on survey data (generally from the Behavioral Risk Factor Surveillance System, or BRFSS). These are 95% confidence intervals, meaning that there is a 95% probability that confidence interval contains the true value.
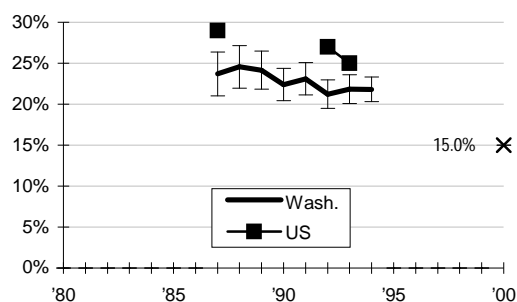
Most confidence intervals in this publication were calculated by multiplying the standard error by 1.96. Standard errors were generated in SUDAAN, a software package used by the CDC and by the Washington Center for Health Statistics for analyzing BRFSS data.

Confidence intervals are sometimes presented in narrative form. For example, in the Tobacco Use and Exposure section there is a statement that, "In 1993, 21.8% (±1.7) of Washington adults reported current smoking." The 1.7 is the figure calculated by multiplying the standard error by 1.96, and it is both added to and subtracted from the observed data point (21.8) to get the 95% confidence intervals. In this case, we have 95% confidence that the true prevalence of current smoking among Washington adults in 1993 was somewhere between 20.1% and 23.5%.

Confidence intervals in this publication are also presented graphically, as in the time trend chart below, showing smoking prevalence from 1987 through 1994. The total intervals are shown by the vertical lines, with the upper and lower limits shown by horizontal lines at each end of the intervals.

Note that the confidence intervals in this chart get smaller in more recent years. This is probably because the size of the sample increased.

Current Smoking Prevalence
Percent of adults who smoke, with Washington year 2000 goal

Confidence intervals in this publication are also presented in some of the horizontal bar graphs, as in the example below showing smoking prevalence in 1994 by age and gender. Rates are significantly different if the confidence intervals do not overlap. For example, there is no significant difference between the rates for men age 65+ and women age 65+, but there is a significant difference between the rates for men age 65+ and men age 35-44.

Smoking Prevalence by Age and Gender
Percent of adults who smoked, Wash. 1994
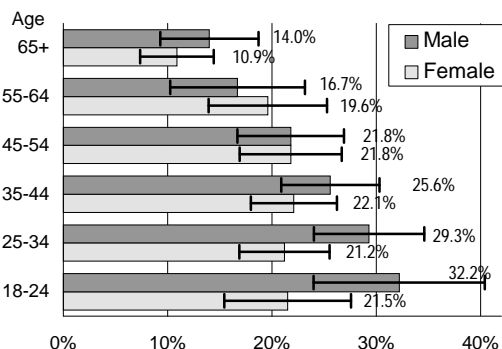
Confidentiality

All of the data in this report are presented in aggregate form, which means that individuals cannot be identified from the data. However, it is important for potential data users to be aware of confidentiality issues related to the data. The medical and health information on birth and fetal death certificates is confidential and may not be linked to any identifying information except for research projects approved by the Human Research Review Board of the Department of

Social and Health Services and the Department of Health. Some death data (particularly causes of death such as suicide and AIDS), while not confidential by law, are extremely sensitive.

It is the responsibility of all data users to release data only in a way such that individuals may not be identified. To ensure continued reporting of statistical information, data must be handled in such a way as to ensure the privacy of individuals and of providers as provided under state law and attendant administrative code.

Congenital anomalies

Congenital anomalies include the following ICD-9 codes: Anencephalus (740), Spina Bifida and Meningocele (741), Other Central Nervous System Anomalies (742), Anomalies of the Heart (745-746), Other Anomalies of Circulatory System (747), Anomalies of the Respiratory System (748), Anomalies of the Digestive System (749-751), Anomalies of the Genitourinary System (752-753), Anomalies of the Musculoskeletal System (754-756), Chromosomal Anomalies (758), Other Congenital Anomalies (759).

Crude rates

A rate is the number of events (such as deaths) in a specified time period divided by the number of people at risk of these events in that period (typically, a state or county population, or the number of births in the case of infant death). This figure is generally multiplied by a constant such as 1,000 or 100,000 to get a number that is easy to read and compare and is reported as 'per 1,000' or 'per 100,000.'

Rates calculated in this manner are called crude rates. They adjust for differences in population size but not differences in population characteristics. These population characteristics also need to be considered in interpreting comparisons. For example, since death rates increase with increasing age, a county with an older population may have higher death rates just because its population is older.

Data Quality

Conclusions and health policy decisions are only as good as the data that go into making them. The highest quality health data are complete, accurate, and timely. They have both precision (are based on accurate counts or statistical estimates) and validity (measure what they are supposed to) and

Studies of validity of reporting systems usually look for an independent source of the information and determine the consistency with data contained in the reporting system.

A factor that affects the completeness of data is the number of unknowns among responses. Sometimes providers do not complete all items. The information may be overlooked or refused by the informant, or the informant may not have been asked for the data. Missing data decrease the overall accuracy of an item because we don't know where they fit (e.g., are smokers less likely to respond to a question on smoking?).

Timely data should be available when needed to be most useful for planning and program assessment purposes. There is often a tradeoff between timeliness and accuracy. For example, if birth certificates are filed quickly, there may not be enough time for malformations or complications (such as fetal alcohol syndrome) to become evident. Similarly, if death certificates are filed quickly, there may not be time for autopsy results to be incorporated into the cause of death data.

Event count

The absolute magnitude of population burden, rather than a rate. Example: There were 312 homicide deaths in Washington in 1994.

Hispanic origin

"Origin" as used by the Census Bureau refers to "the ancestry, nationality group, lineage, or country of birth of the person or the person's parents or ancestors before their arrival in the United States." Persons of Hispanic origin have their origins in a Hispanic or Spanish-speaking country such as Mexico, Cuba, or Puerto Rico, or the Spanish-speaking countries of Central or South America. Persons of Hispanic origin may be of any race.

Hispanic origin was added as an ethnic category in the vital records system and collected as a separate item (in addition to race) in 1988. Prior to 1988, Hispanic data were provided by a racial category of "Mexican/Chicano" or "Mexican American". Data based on Hispanic

origin from those years are not comparable with data collected in 1988 or thereafter and should not be used for trend analysis or year to year comparisons.

The certificates for live births, deaths, and fetal deaths in Washington State capture Hispanic origin under two separate items, one to measure ethnicity and another to measure race. The item measuring ethnicity asks, [Is the person] "Of Hispanic Origin or descent (Ancestry)?" and permits a "Yes/No" response. The item measuring race on the birth and fetal death certificates says, "Race (American Indian, White, Black, Asian/Pacific Islander (Specify subgroup), etc.). On the death certificate, the item reads, "Race (Specify)." Although Hispanic is not listed as a sample response under "Race" beginning in 1992, it is still provided as a racial designation, particularly for people who have already been classified as being of Hispanic origin ethnicity.

The National Center for Health Statistics (NCHS), however, does not treat Hispanic Origin as a race and requires instead that persons reporting Hispanic as a race be counted as "White."

In a few instances, the Hispanic ethnicity question was marked unknown, yet "Hispanic" was given as the person's race. Beginning in 1992, if a person's ethnicity was marked "Unknown" and his/her race was given as "Hispanic," then that person's ethnicity was counted as Hispanic Origin. Only a small number--about ten births and 60 deaths each year--were reclassified in this way. While this change only causes a 0.1% difference in the number of Hispanics at birth, it increases the number of Hispanics at death by 14%.

During the production of this document, some reviewers noted that, for many causes of death (including total deaths, coronary heart disease, stroke, most cancers, and chronic obstructive pulmonary disease) the death rates shown for Hispanics were considerably lower than those for non-Hispanics. This prompted some examination of whether this phenomenon is real or accounted for by some anomaly such as undercounting of people of Hispanic origin in the death system.

Based on linked birth-death files and correspondence of Washington data with national data, it appears that the lower Hispanic death rates are probably real, particularly in older age groups. There is some evidence that among younger Hispanics, death rate patterns are closer to those of the general population; in fact, for some causes the rates seem to be higher.

Infant mortality rates

The period infant mortality rate per 1,000 live births is computed by dividing the number of infant deaths for a given year by the number of live births, then multiplying by 1,000. This differs from a cohort of infants which would include infants born in one year who died within one year of birth, even if they died in the next year.

Low Birth Weight

Traditionally, low birth weight has been defined as 2,500 grams or less. However, the International Classification of Diseases, Ninth Revision (ICD 9) redefines low birth weight as less than 2,500 grams. Thus, according to national guidelines, a birth weight equal to 2,500 grams would now be included with the normal birth weight category. As of 1994, the birth weight categories published in Washington's vital statistics reports have been revised, in accordance with these guidelines.

PUMS

See "Small Numbers."

Population Data (denominators)

Population data presented or used to calculate rates in this report are from the decennial U.S. Census or are intercensal year estimates provided by the Washington State Office of Financial Management, Forecasting Division.

Prenatal care considerations

Birth certificate limitations. The birth certificate is a useful surveillance tool to determine trends in pregnancy outcomes, but the scope of prenatal behavioral information is limited. The birth certificate provides limited data on maternal behavior during pregnancy and no information on preconception behavior. It also provides no information regarding the timing of prenatal care visits. The validity of

prenatal care information recorded on the birth certificate is limited.

Selection bias. Women who receive adequate prenatal care are a self-selected group and differ greatly from those who receive inadequate care. Selection bias exists in all evaluations of the association between prenatal care and pregnancy outcomes. The ways in which women who receive adequate care and those who do not differ are not fully known. Even if basic social and demographic variables are controlled for, the two groups will probably differ in respects that have not been measured. ("Prenatal Care", John L. Kiely, PhD, et al, *From Data to Action: CDC Public Health Surveillance of Women, Infants, Children.* USDHHS, no date.)

Prenatal Care Indices. The Kessner Index of Prenatal Care uses trimester that care began in classifying the adequacy of care received. The Kotelchuck Index of Prenatal Care uses month that care began in classifying the adequacy of care received. Kessner's Index is truncated at gestational age 36 with a minimum number of visits of 9. The Kotelchuck Index adjusts the minimum number of visits for the month that care began and creates a continuous scale consisting of the ratio of the observed number of visits to the adjusted expected number of visits.

Race and Ethnicity

Race data collected on vital statistics follow the definition established by the Census Bureau, as follows:

"The concept of race as used by the Census Bureau reflects self-identification; it does not denote any clear-cut scientific definition of biological stock. The data for race represents self-classification by people according to the race with which they most closely identify. Furthermore, it is recognized that the categories of the race item include both racial and national origin or socio-cultural groups."[1]

Birth and death certificates use open-ended reporting of race, allowing for multiple racial entries. The determination of race for multiple

racial entries follows decision rules established by the National Center for Health Statistics (NCHS): If more than one race is listed and one is Hawaiian, then Hawaiian is selected. For other races, the first race listed is selected. Prior to 1990, the first non-white race listed was selected (unless Hawaiian was listed somewhere in the item). This change increases the number of events coded 'White' and decreases the number of events coded to other races. Thus, race-specific birth or death rates may not be comparable for the two time periods before and after the change.

To determine race of newborns on birth certificates, each parent's race is first established according to the previous rule. Rules to determine race of the child, then, fall along a similar pattern:

· If either parent is Hawaiian, then the child's race is considered Hawaiian.

· In cases where one parent is white and the other non white, the non white race is selected as the race of the child.

· If both parents are non white, the father's race is selected except where the mother's race is Hawaiian.

Child's race data in Washington's vital statistics system and reports use the calculated child's race. In addition to this calculated race, the mother also has the choice of designating a race/Hispanic origin for her child. These designated race data are only available starting in 1992.

Reporting of race/Hispanic origin on death certificates is sometimes based on observing the decedent, rather than questioning the next of kin. This procedure causes an underestimate of deaths for certain groups, particularly Native Americans, some of the Asian subgroups, and Hispanics. Thus, death rates based on death certificate data are lower than true death rates for these groups.

For infant deaths, the infant's race on the death certificate is related to the race of the infant's parents. If both parents belong to a particular non-white racial group, the child is much more likely to be given that particular race than if only one parent belongs to that racial group and the other parent is white. For this reason, the most accurate race-specific infant mortality rates come from linked birth-infant

---

[1] U.S. Bureau of the Census, *1990 Census of Population: General Population Characteristics, Washington,* Report 1990 CP-1-49, Washington, D.C., June 1992.

death data files, where the child's race at birth can be used for both the numerator and the denominator.

(See also "Hispanic origin")

Small numbers

Many of the numbers and rates reported in this document are based on vital statistics such as births and deaths and can be assumed to be very close to actual numbers and rates. Others, such as those for many risk behaviors such as smoking, are based on samples of the population.

We do know the actual number of births, deaths, and population (assuming complete reporting of events), so we can calculate an exact birth or death rate for any one year. However, the data may still be affected by random fluctuations in the number of events between successive measurements (e.g., for different years).

The effect of such random fluctuations on birth or death rates is proportionately larger when the number of events is small. For example, one more infant death has a larger numerical impact on an area with 3 deaths than it does on an area with 300 deaths. Because of these random fluctuations, the rates based on small numbers may not be as stable as those based on larger numbers in the sense that they may have limited predictive value. Specifically, knowing one year's rate in such instances may not allow one to reliably anticipate the rate for another year. This instability makes it difficult to use the rates for program planning or assessment purposes. In fact, considerable caution should be used in interpreting any data where the number of events is small.

In the presentation of survey data by county in this publication, contiguous counties with small populations are combined to form geographic regions large enough for statistically reliable analyses. This is done according to a system defined by the U.S. Bureau of the Census and adopted by the Washington State Office of Financial Management. The regions are referred to as PUMS regions (from Public Use Microdata Sample, the term used by the Census Bureau to refer to the files containing records of data on these regions). The largest counties stand alone in this analysis.

In this document, rates from surveys are not presented if they are based on five or fewer events. In such cases the graphic presentation uses the notation "NR" which means "not reliable." These rates are suppressed because they are subject to misinterpretation and quite likely to change dramatically from one time period to another.